Adaptive Estimation of Hidden Semi-Markov chains with Parameterised Transition Probabilities and exponentially decaying states *

Jason Ford † Vikram Krishnamurthy † John B. Moore †

Abstract

This paper develops maximum likelihood (ML) estimation schemes for finite-state semi-Markov chains in white Gaussian noise. We assume that the semi-Markov chain is characterised by transition probabilities of known parametric form with unknown parameters. We reformulate this hidden semi-Markov model (HSM) problem in the scalar case as a two-vector homogeneous hidden Markov model (HMM) problem in which the state consist of the signal augmented by the time to last transition. With this reformulation we apply the expectation Maximumisation (EM) algorithm to obtain ML estimates of the transition probability parameters, Markov state levels and noise variance. To demonstrate our proposed schemes, motivated by neuro-biological applications, we use a damped sinusoidal parameterised function for the transition probabilities.

1 Introduction

Hidden Markov models (HMMs) with homogeneous (time invariant) transition probabilities have been widely used in communication system, speech processing and biological signal processing. However, in some neuro-biological systems and computer communication networks the transition probabilities of the Markov chain are a function of the time to last transition. That is, the transition probabilities depend on the time the Markov chain has spent in a particular state. Such inhomogeneous Markov chains are termed semi-Markov chains.

In this paper we propose maximum likelihood (ML) estimation schemes for finite-state semi-Markov chains in white Gaussian noise. We assume that the semi-Markov chain is characterised by transition probabilities of known parametric form with unknown parameters. We reformulate this hidden semi-Markov model (HSM) problem in the scalar case as a two-vector homogeneous hidden Markov model (HMM) problem in which the state consist of the signal augmented by the time to last transition. With this reformulation we apply the expectation Maximumisation (EM) algorithm to obtain ML estimates of the transition probability parameters, Markov state levels and noise variance. It is theoretically possible in some semi-Markov models that the number of possible discrete state for the time to last transition is the number of observation in the data set. However, it turns out that in many models with practical significance (including DEDS models which we describe below) the number of quantisation levels for this time variable can be significantly reduced via aggregation and thus leads to substantial reduction in computational requirements.

To demonstrate our proposed schemes, motivated by neuro-biological applications, we consider one example of a parameterised function for the transition probabilities. This consists of a damped exponentially decaying sinusoidal (DEDS) function with unknown parameters including amplitude, phase, frequency and decay rate. This signal model is used in the bio-phyiscal literature to model ion channel currents in cell membranes.

The paper is organised as follows: In Section 2 we describe a semi-Markov Model. In section 3 we show how the general semi-Markov model from Section 2 can be reformulated so as to be equivalent to standard Markov model. The most important condition for equivalent will be shown to be the aggregation of states. When a semi-Markov model is then hidden in noise we call it a hidden semi-Markov model. The standard hidden semi-Markov problem is formulated here. In Section 4, we develop the re-estimation formulae for the HSMM problem described in Section 3. In Section 5, we apply a particular function to the previous re-estimation formulae and demonstrate the learning characteristic of the adaptive estimation.

2 Problem Formulation

In this section first the signal model is described. Next we reformulate the scalar semi-Markov Hidden Markov models as augmented 2-vector Hidden Markov Models. To do this we develop the idea of state aggregation.

2.1 Signal Model

We first describe a general scalar finite-state semi-Markov model. Then an example of one such model called Damped exponentially Decaying sinusoidal (DEDS) which has neurobiological applications is given. ADD A bit about decaying states and aggregation

Multi-state semi-Markov model

Consider a discrete-time, finite-state stochastic process s_k, $k \geq 0$, where for each k, s_k is a random variable taking on a finite number N_s of possible states $q_1, \dots, q_{N_s}$. Assume that the parameterised transition probabilities $a^{\theta}_{ij}(k)$, parameterised by θ, are defined at time k as $a^{\theta}_{ij}(k) \triangleq P(s_{k+1} = q_j | s_k = q_i)$. In a semi-Markov model $a^{\theta}_{ij}(k)$ is a function of the time to the last transition at time k, denoted t_k, and so is a function of the form:

$$a^{\theta}_{ij}(t_k) : t_k \mapsto [0,1] \quad \text{for } i,j \in [1, N_s] \tag{2.1}$$

since a_{ij} are probabilities.

*Partially supported by D.S.T.O Australia and Boeing (BCAC), U.S.A.

†Department of Systems Engineering, Research School of Physical Sciences and Engineering, Australian National University, GPO Box 4, Canberra ACT 2601.

For convenience we work with the number of discrete-time samples after the last transition rather than the actual time. Thus for a sampling time interval of T_s, t_k is quantised to a finite number N_t of possible integer values $\tau_1, \dots, \tau_{N_t}$, such as $\tau_i = i$ or $\tau_i = 2i$. Without any computational effort and memory constraints, it would be reasonable to take τ_i as the integer i, for $i = 1, \dots, N_t$. An upper bound on N_t is the total number of observations in the data set. However, as described later, state aggregation allows N_t to be bounded by more realistic values. Notice that when $a_{ij}^{\theta}(t_k)$ is independent of t_k, the transition probabilities are independent of time, and s_k reduces to a homogeneous first order Markov process.

DEDS example

In our case the number of physical states is limited to $N_s = 2$ and the transition probabilities vary as

$$
\begin{aligned}
a_{ii}^{\theta}(t_k) &= \exp(-\tau_i t_k)(d_i + b_i \sin(\omega_i t_k + \phi_i)) + c_i \\
a_{ij}^{\theta}(t_k) &= 1 - a_{ii}^{\theta}(t_k)
\end{aligned}
\tag{2.2}
$$

where $\theta = \theta_i$ and $\theta_i = (\tau_i, d_i, b_i, \omega_i, \phi_i, c_i)$.

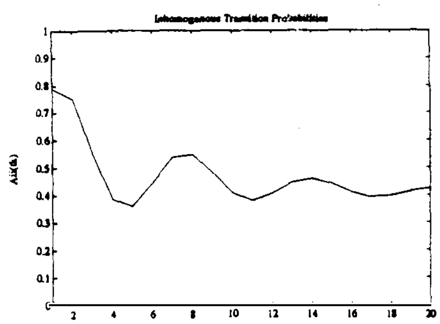


Figure 2.1: $r = 0.3, d = .3, b = .4, \omega = 1, \phi = 0, c = .4$

Exponentially Decaying States

Here the state levels decay as the chain remains in state. That is, we assume that transitions occur at time instants $\Delta, 2\Delta, \dots, l\Delta, (l+1)\Delta, \dots$. Also the state decays over the intervals $[l\Delta + 1, (l+1)\Delta]$ as follows

$$
\text{If } s_{l\Delta+1} = q_i \text{ then } s_k = \rho^{k-(l\Delta+1)} q_i, \quad k \in [l\Delta+1, (l+1)\Delta]
\tag{2.3}
$$

2.2 Formulation of semi-Markov process as a vector homogeneous Markov process

The class of scalar semi-Markov models in (2.1) can be modelled as a homogeneous first order 2-vector Markov process as follows: Define the 2-vector process $\mathbf{S}_k$ as $\mathbf{S}_k = (s_k, t_k)$ for each $k \geq 0$. Clearly $\mathbf{S}_k$ is a finite-state process with $N = N_s N_t$ states. Here N_t is taken as the maximum duration time in any state considering the observation sequence of length T.

It is easily shown that the 2-vector stochastic process $\mathbf{S}_k$, as defined above with $\tau_i = i$ for $i = 1, \dots, N_t$ is a homogeneous, first order Markov process (see [7]).

Notice that

$$
t_{k+1} = \begin{cases} t_k + 1 & \text{if } s_{k+1} = s_k \text{ and } t_k < N_t \\ 1 & \text{otherwise} \end{cases}
\tag{2.4}
$$

So t_{k+1} depends only on t_k, s_k and s_{k+1}. Also from (2.4) the transition probabilities for the homogeneous vector process $\mathbf{S}_k$ are for $1 \leq \tau_h \leq N_t$:

$$
P(\mathbf{S}_{k+1} | \mathbf{S}_k) = \begin{cases} a_{ii}(\tau_h)^{\theta} & \text{if } \mathbf{S}_k = (q-i, \tau_h) \\ & \quad \text{and } \mathbf{S}_{k+1} = (q_i, \tau_h + 1) \\ a_{ij}(\tau_h)^{\theta} & \text{if } \mathbf{S}_k = (q_i, \tau_h) \\ & \quad \text{and } \mathbf{S}_{k+1} = (q_i, 1),\ i \neq j \\ 0 & \text{otherwise} \end{cases}
\tag{2.5}
$$

which are independent of k.

Remark: If for some integer $N_t' < N_t$, τ_i is defined more generally as $\tau_i = i$, $i = 1, \dots, N_t' - 1$ and $\tau_{N_t'} = i : N_t' \leq i \leq N_t$, the first order Markov property still holds. However, it is readily shown that $\mathbf{S}_k$ is then not necessarily homogeneous. For certain functions $a_{ij}(t_k)^{\theta}$ that "saturate" beyond some t_k, $\mathbf{S}_k$ is homogeneous and aggregation of states in the saturation region is possible as discussed below. ◦

Notation: Denote the set of $N = N_s N_t$ states $\{(q_1, \tau_1), \dots, (q_{N_s}, \tau_{N_t})\}$ as $\{\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_N\}$, although, not necessarily in the same order. We will denote eleme of this set by integer subscripts, usually m or n. A for $Q_m = (q_i, \tau_h)$ and $Q_n = (q_j, \tau_l)$ where $m, n \in [1,$. $\tau_h, \tau_l \in [1, N_t]$, $q_i, q_j \in [1, N_s]$, denote the transition pro bilities of the homogeneous $\mathbf{S}_k$ process by

$$
\begin{aligned}
\mathbf{A}^{\theta} &= (a_{mn}^{\theta}) \\
a_{mn}^{\theta} &= a_{(h,i),(l,j)}^{\theta} \triangleq P(\mathbf{S}_{k+1} = \mathbf{Q}_n | \mathbf{S}_k = \mathbf{Q}_m). \quad (2.
\end{aligned}
$$

2.3 Aggregation

In proving that $\mathbf{S}_k$ is a homogeneous Markov process abov it is assumed that $\tau_i = i$ and N_t is the maximum duratio time in any state. Since we may not know this value we ca set $N_t = T$, the length of the entire observation sequence but then the number of states $N = N_t N_s$ will be excessiv for computational purposes. Is it possible to quantise the time to the last transition to N_t' states, where $N_t' << T$ with negligible error? We propose to do so by "aggregating" [11] the states $(q_i, \tau_{N_t'}), (q_i, \tau_{N_t'+1}), \dots, (q_i, \tau_{N_t})$ into a aggregated state

$$
\overline{\mathbf{Q}}_{N_t', i} = \{(q_i, \tau_{N_t'}), (q_i, \tau_{N_t'+1}), \dots, (q_i, \tau_{N_t})\}.
\tag{2.7}
$$

Aggregation Property: We have proved in [7] that the above aggregation leads to negngible small errors for a certain class of functions $a_{ij}(t_k)$ defined in (2.1) satisfying for arbitrary small $\epsilon > 0$ and $N_t' < t_k < N_t = T$

$$
|a_{ij}^{\theta}(t_k) - a_{ij}^{\theta}(t_{k-1})| < \epsilon \quad \text{for } i, j \in [1, N_s]
\tag{2.8}
$$

More specifically, the DEDS model with transition probabilities that exponentially converge, and the discrete semi-Markov process with Poisson distributed transition times satisfy (2.8) and so can be aggregated. Simulations show that in most cases for such processes, choosing N_t so that $\epsilon < 0.05$ is adequate. For the rest of this paper we set $N_t = N_t' << T$ where N_t' is suitably large (usually less than 50) to result in negligible error. Of course, now (2.4) is modified as

$$t_{k+1} = \begin{cases} t_k + 1 & \text{if } s_{k+1} = s_k \text{ and } t_k < N_t \\ t_k & \text{if } s_{k+1} = s_k \text{ and } t_k = N_t \\ 1 & \text{otherwise} \end{cases} \quad (2.9)$$

Also from (2.4) we have

$$P(\mathbf{S}_{k+1}|\mathbf{S}_k) = \begin{cases} a^{\theta}_{ii}(\tau_h) & \text{if } \mathbf{S}_k = (q_i, \tau_h) \text{ and } \\ & \mathbf{S}_{k+1} = (q_i, \tau_h);\ \tau_h = N_t \\ a^{\theta}_{ii}(\tau_h) & \text{if } \mathbf{S}_k = (q_i, \tau_h) \text{ and } \\ & \mathbf{S}_{k+1} = (q_i, \tau_h + 1);\ 1 \le \tau_h < N_t \\ a^{\theta}_{ij}(\tau_h) & \text{if } \mathbf{S}_k = (q_i, \tau_h) \text{ and } \\ & \mathbf{S}_{k+1} = (1, q_j),\ i \neq j,\ 1 \le \tau_h \le N_t \\ 0 & \text{otherwise} \end{cases} \quad (2.10)$$

Hence the transition probability from state $\mathbf{S}_k = (t_k, s_k)$ depends only on the previous state $\mathbf{S}_{k-1} = (t_{k-1}, s_{k-1})$. Another approach of aggregation, which assumes a periodic transition probability, would be to aggregate states together that occurred at the same part of a period. In this case the transition probabilities would not be required to decay to a constant. The corresponding aggregation error is given by

$$|a_{ij}(t_k + P) - a_{ij}(t_k)| < \epsilon \quad (2.11)$$

where $\mathbf{P}$ is the calculated period.

Transition probability matrix sparseness: Notice that with $\tau_h = h$ for $h < N_t = N'_t$, the relationship between (2.1) and (2.6) is

$$a^{\theta}_{ii}(\tau_h) = a^{\theta}_{(h,i),(h+1,i)}, \text{ and } a^{\theta}_{ij}(\tau_h) = a^{\theta}_{(h,i),(1,j)}, \quad i \neq j,\ 1 \le \tau_h \le N_t. \quad (2.12)$$

Clearly $\mathbf{A}^{\theta}$ has $(N_s N_t)^2 = N^2$ elements. However, since t_{k+1} is restricted as in (2.9) to only three possible values, simple calculations show that $(N^2 - N_s^2 N_t)$ elements of $\mathbf{A}^{\theta}$ are zero. For $i, j \in [1, \ldots N_s]$, only the following elements of $\mathbf{A}^{\theta}$ are not necessarily zero:

$$a^{\theta}_{(h,i),(h+1,i)}, \quad \tau_h \neq N_t; \quad a^{\theta}_{(h,i),(1,j)}, \quad i \neq j; \quad \text{and} \quad a^{\theta}_{(N_t,i),(N_t,i)}. \quad (2.13)$$

Consequently, in any scheme to estimate $\mathbf{A}^{\theta}$, only the $N_s^2 N_t$ elements of A^{θ} in (2.13) need be estimated.

2.4 Hidden semi-Markov Models

Semi-Markov processes embedded in noise are called Hidden Semi-Markov Models (HSMMs). Given the semi-Markov process $\mathbf{S}_k = (s_k, t_k)$ hidden are, that is indirectly observed by measurements y_k. We denote the sequence the sequence $y_1, y_2, \ldots, y_k$ by $\mathbf{Y}_k$. The vector of probability functions $\mathbf{b}(.) = (b_m(.)) = P(y_k|\mathbf{S}_k = \mathbf{Q}_m)$ where $\mathbf{Q}_m = (q_i, \tau_h)$ is assumed invariant of the times k and τ_h. So $b_m(y_k) = b_i(y_k) = f(y_k|s_k = q_i)$. Also assume the independence property $f(y_k|s_k = q_i, s_{k-1} = q_j, \mathbf{Y}_{k-1}) = f(y_k|s_k = q_i)$. Assuming corrupted by zero mean, normally distributed white noise as follows: $y_k = s_k + w_k$, $w_k \sim N[0, \sigma_w^2]$. Then $b_i(y_k) = (\sqrt{2\pi}\sigma_w)^{-1} exp(|y_k - q_i|^2/(2\sigma_w^2))$.

Further, assume that the initial state probability vector $\underline{\pi} = (\pi_m)$ is defined from $\pi_m = P(\mathbf{S}_1 = \mathbf{Q}_m)$. The transition probabilities $\mathbf{A}^{\theta}$ are defined as in [6]. The vector HMM for the $\mathbf{S}_k$ process is denoted $\lambda = (\mathbf{A}^{\theta}, \mathbf{b}(.), \underline{\pi})$. Of course λ also denotes the HSMM for the $\mathbf{S}_k$ process.

3 Estimation and Maximumisation of HSMs the EM algorithm

3.1 Estimation Objectives

Given the Semi-Markov model described above which has been formulated as a Markov model, and a observation sequence $y_1, y_2, y_3, \ldots, y_k$ denoted $\mathbf{Y}_k$ there are three standard HHM problems which can be solved.

State Estimation: Given a noisy observation sequence Y_T how do we obtain a maximum a posteriori (MAP) state estimates of the semi-Markov chain, $\hat{\mathbf{S}}_k$ $k \in [1, \ldots, T]$.

Parameter Estimation: Find the ML estimate λ^{ml} of the HSM, where $\lambda^{ml} = argmax_{\lambda}\ f(Y_t|\lambda)$.

3.2 The EM algorithm

We use the EM algorithm [9] as follows. There are two steps to the EM algorithm. The E-step and the M-step. The E-step involves the calculation of Q-function. *Definition:* The Q-function is defined as

$$Q(\lambda, \overline{\lambda}) = \sum_{Q} P(Q|O, \lambda) \log[P(O, Q|\overline{\lambda})] \quad (3.14)$$

The M-step requires the maximumisation of the Q-function over the parameter space.

E-step

The E-step requires the calculation of the Q-function, *Definition:* The Q-function is defined as

$$Q(\lambda, \overline{\lambda}) = \sum_{Q} P(Q|O, \lambda) \log[P(O, Q|\overline{\lambda})]$$

$$= \sum_{k=1}^{T} \sum_{m=1}^{N} \sum_{n=1}^{N} \xi(m, n) \log a^{\theta}_{mn} + \sum_{k=1}^{T} \sum_{n=1}^{N} \gamma_k(n) \log b_n(q_n)$$

$$\text{where } b_n(q_n) = \frac{1}{\sigma_w} \exp\left(\frac{-(y_n - q_n)^2}{2\sigma_w^2}\right)$$

$$\text{and } \gamma_k(n) = \frac{\alpha_k(n)\,\beta_k(n)}{\sum_{m=1}^{N} \alpha_k(m)\,\beta_k(m)}$$

$$\text{and} \quad (m, n) = \frac{\alpha_k(m)\, a_{mn}\, b_n(y_{k+1})\,\beta_{k+1}(n)}{\sum_{g=1}^{N}\sum_{h=1}^{N} \alpha_k(g)\, a_{gh}\, b_h(y_{k+1})\,\beta_{k+1}(h)} \quad (3.15)$$

To calculate the α, β, γ and ξ variables the forward-backward procedure can be used. **Forward-backward Procedure** As with the standard HMM problem recursive formula for updating the forward variable α_k and then backward variable are readily found.

$$\pi_n b_n(y_1)$$

$$\left(\sum_{m=1}^{N} \alpha_{k-1}(m)\, a_{mn}\right)$$

$$\beta_T(n) \quad 1$$

$$\sum_{m=1}^{N} a_{nm}\, b_m(y_{k+1})\, \beta_{k+1}(m).$$

where n,m are the vector states

(3.16)

[illegible] and β_k requires of the order $N_s^2 N_t^2 T$ floating [illegible] operations although a lot of the transition probabil[illegible] could be zero and with little refinement the number of [illegible] can be reduced to $N_s^2 N_t T$. The varialbes can now [illegible] to calculate the Q-function.

[illegible]

[illegible] the Baum-Welch re-estimation formulae requires the [illegible] derivatives of the auxiliary function, Q. For par[illegible] functional forms of the transition probabilities the [illegible] derivatives will differ. The general procedure used [illegible] update the estimates the parameters is to find the ze[illegible] of the partial derivatives and then use these zeros as [illegible] estimations of the parameters. That is,

$$\text{for the (l+1)th pass,} \quad \hat{\theta_{l+1}} = \theta_l : \frac{\partial Q_l}{\partial \theta_l} = 0. \tag{3.17}$$

[illegible] estimation formulas are based on a single update at [illegible] pass. However, for parametric forms with numerous [illegible] parameters it may be possible to update several parameters [illegible] each pass under some conditions.

$$\hat{q_n} = \frac{\sum_{k=1}^{T} \gamma_k(n)\, q_n}{\sum_{k=1}^{T} \gamma_k(n)} \tag{3.18}$$

$$\hat{\sigma_w^2} = \frac{1}{T} \sum_{k=1}^{T} (y_k - \sum_{n=1}^{N} \gamma_k(n)\, q_n)^2 \tag{3.19}$$

4 Application

[illegible] two example of the application of the EM algorithm [illegible] given.

4.1 The DEDS example

To demonstrate our proposed scheme we use one example [illegible] parameterised function which has been motivated by a par[illegible] ticular biological signal problem. We define the transition probabilities in terms of the allowable physical states $\mathbf{q} =$ [illegible] q_N, and the time from last transition $t_k = 1, 2, \ldots N_t$, where these are measured in terms of sampling periods rather then any actual time unit. Remembering that the semi-Markov state is represented as $\mathbf{S}_k = (q_i, t_k)$. In our [illegible] the number of physical states is limited to $N_s = 2$ and [illegible] transition probabilities vary as

$$\begin{aligned} a_{ii}(t_k) &= \exp(-r_i t_k)(d_i + b_i \sin(\omega_i t_k + \phi_i)) + c_i & \text{if } t_k < N_t \\ a_{ii}(t_k) &= \exp(-r_i N_t)(d_i + b_i \sin(\omega_i N_t + \phi_i)) + c_i & \text{if } t_k \geq N_t \\ a_{ij}(t_k) &= 1 - a_{ii}(t_k) & \text{if } i \neq j \end{aligned} \tag{4.20}$$

[illegible] assuming that the state levels and noise variance are known, because these problem have little interest, Q becomes

$$\begin{aligned} Q = &\sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(1,t,1,t+1) \log[e^{-r_1 t}(d_1 + b_1 \sin(\omega_1 t + \phi_1)) + c_1] \\ &+ \sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(1,t,2,t+1) \log[1 - e^{-r_1 t}(d_1 + b_1 \sin(\omega_1 t + \phi_1)) - c_1] \\ &+ \sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(2,t,2,t+1) \log[e^{-r_2 t}(d_2 + b_2 \sin(\omega_2 t + \phi_2)) + c_2] \\ &+ \sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(2,t,1,t+1) \log[1 - e^{-r_2 t}(d_2 + b_2 \sin(\omega_2 t + \phi_2)) - c_2] \end{aligned} \tag{4.21}$$

To maximise this unconstrained function, noting that the constraint that probabilities must sum to one has already been included in the function, we must equate all the partial derivatives of Q, with respect to the model parameters, to zero.
These partial derivatives are

$$\begin{aligned} \frac{\partial Q}{\partial r_i} = &\sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(i,t,i,t+1) \frac{-r_i e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i))}{e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i)) + c_i} \\ &+ \sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(i,t,j,t+1) \frac{-r_i e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i))}{e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i)) + c_i - 1} \end{aligned} \tag{4.22}$$

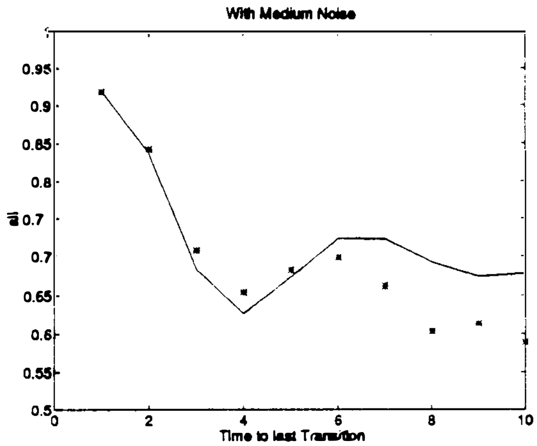
$$\begin{aligned} \frac{\partial Q}{\partial b_i} = &\sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(i,t,i,t+1) \frac{e^{-r_i t} \sin(\omega_i t + \phi_i)}{e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i)) + c_i} \\ &+ \sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(i,t,j,t+1) \frac{e^{-r_i t} \sin(\omega_i t + \phi_i)}{e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i)) + c_i - 1} \end{aligned} \tag{4.23}$$

$$\begin{aligned} \frac{\partial Q}{\partial c_i} = &\sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(i,t,i,t+1) \frac{1}{e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i)) + c_i} \\ &+ \sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(i,t,j,t+1) \frac{1}{e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i)) + c_i - 1} \end{aligned} \tag{4.24}$$

$$\begin{aligned} \frac{\partial Q}{\partial d_i} = &\sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(i,t,i,t+1) \frac{e^{-r_i t}}{e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i)) + c_i} \\ &+ \sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(i,t,j,t+1) \frac{e^{-r_i t}}{e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i)) + c_i - 1} \end{aligned} \tag{4.25}$$

$$\begin{aligned} \frac{\partial Q}{\partial \omega_i} = &\sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(i,t,i,t+1) \frac{e^{-r_i t} b_i t \cos(\omega_i t + \phi_i) \sin(\omega_i t + \phi_i)}{e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i)) + c_i} \\ &+ \sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(i,t,j,t+1) \frac{e^{-r_i t} b_i t \cos(\omega_i t + \phi_i) \sin(\omega_i t + \phi_i)}{e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i)) + c_i - 1} \end{aligned} \tag{4.26}$$

$$\frac{\partial Q}{\partial \phi_i} = \sum_{k=1}^{T} \sum_{t=1}^{N_t} \xi_k(i,t,i,t+1) \frac{e^{-r_i t} b_i \cos(\omega_i t + \phi_i) \sin(\omega_i t + \phi_i)}{e^{-r_i t}(d_i + b_i \sin(\omega_i t + \phi_i)) + c_i}$$



r=0.15,b=0.9,c=.6,d=.4,ω=1.25 and ϕ=0

Figure 4.1: The estimation of semi-Markov chain in medium noise

$$+\sum_{k=1}^{T}\sum_{t=1}^{N_t}\xi_k(i,t,j,t+1)\frac{e^{-r_it}b_i\cos(\omega_it+\phi_i)\sin(\omega_it+\phi_i)}{e^{-r_it}(d_i+b_i\sin(\omega_it+\phi_i))+c_i-1} \quad (4.27)$$

There are no analytic solutions for the zeros of these partial derivatives, so only numerical solutions can be obtained. *NOTE* : It was not simply a matter of finding the zero of this partial derivative because these functions generally have multiple zeros, therefore a comparison of all these zeros to determine the maximising value is required to find a re-estimate for the parameter. This also means that a particular algorithm doesn't demonstrate optimal convergence purely because it has found a zero for the partial derivative. The maximising zero must be found hence making the zero-locating algorithm in important part of the algorithm.

4.2 The decaying states

$$\begin{aligned} b_i(\bar{Y}_l) &= \sum_{k=1}^{\Delta}\log b_i(y_{k+l\Delta}) \\ &= \sum_{k=1}^{\Delta}\frac{(y_{k+l}\Delta-\rho^{k-1}q_i)^2}{2\,\sigma_w^2} \end{aligned} \quad (4.28)$$

Hence the Q-function from (3.13) becomes

$$\begin{aligned} Q = & \sum_{n=1}^{T/\Delta}\sum_{i=1}^{N_s}\sum_{j=1}^{N_s}\xi_n(i,j)\log a_{ij} \\ & + \sum_{n=1}^{T/\Delta}\sum_{i=1}^{N_s}\gamma_k(i)\log(\frac{1}{\sqrt{2\,\pi}\sigma_w})\sum_{k=1}^{\Delta}\frac{(y_{k+n\Delta}-\rho_i^{k-1}\,q_i)^2}{2\,\sigma_w^2} \end{aligned} \quad (4.29)$$

4.3 Computer Studies

HSM data: We consider 2-state process ($N_s = 2$). The figure shows the results of simulations run under medium level noise (fig. 4.1). Here the standard deviation of the noise is half that of the difference in levels.

References

[1] B. Sakmann and E. Naher ed., *Single Channel Recording*, Plenum Press, New York, 1983.

[2] S.H. Chung, J.B. Moore, L. Xia, L.S. Premkumar and P.W. Gage, *Hidden Markov Model Techniques for extracting small Ionic currents from noise*, Phil. Trans. of Royal Society, to appear.

[3] V. Krishnamurthy, J.B. Moore, S.H. Chung, *On Hidden Fractal Model Signal Processing*,

[4] L.E. Baum and T. Petrie, *Statistical inference for probabilistic functions of finite state Markov chains*, Ann.Math.Stat., Vol.37, pp 1554-1563, 1966.

[5] L.E. Baum, T. Petrie, G. Soules and N. Weiss, *A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains*, Ann.Math.Stat., Vol.41, No.1, pp 164-171, 1970.

[6] L.R. Rabiner, *A tutorial on Hidden Markov Models and selected applications in speech recognition*, Proc. IEEE, Vol.77, No.2, pp 257-285, 1989.

[7] V. Krishnamurthy, J.B. Moore and S.H. Chung, *Hidden Fractal Model Signal Processing*, Signal Processing, pp.177-192, Vol.24, No.2, Aug.1991.

[8] S.E. Levinson, L.R. Rabiner and M.M. Sondhi, *An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition*, The Bell System Theor. J., No.6, pp 1035-1074, 1983.

[9] A.P. Dempster, N.M Laird, D.B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, J. Royal Stat. Soc., ser 39, vol. 6, pp 1-38, 1977.

[10] D.R. Cox, *Renewal Theory*, Science Paperbacks, London, 1962.

[11] J.G. Kemeny and J.L. Snell, *Finite Markov Chains*, Van Nostrand, N.Y., 1960.